A. Data Preprocessing

All datasets used in the experiments are publicly available. The links to access each dataset are summarized in Table A.2. The data preprocessing of UT Zappos 50k, WIDER Attribute, CUB-200-2011, and ImageNet-100 follows the procedure described in [4], while the data preprocessing of Fitzpatrick17k follows the MEDFAIR framework [7]. The procedures are detailed below for completeness.

UT Zappos 50k. The images are resized to 32 x 32 pixels and randomly split into training and testing sets with a 70/30 proportion. To accommodate the small image size, the first 7x7 kernel of ResNet-50 encoder is replaced with a 3x3 kernel and its first max pooling layer is removed.

WIDER Attribute. The images are resized to 224 pixels along the shorter side using bilinear interpolation. The dataset is annotated with bounding box-specific attributes. As in [4], the bounding box attributes of each image are merged into an image-level attribute vector via the OR operation. In the experiments, the original training and validation sets are combined into one training set.

CUB-200-2011. The images are resized to 224 pixels along the shorter side using bilinear interpolation. The original training and testing sets are used.

Fitzpatrick17k. The images are resized to 256 x 256 pixels and randomly split into training/validation/testing sets with a 80/10/10 proportion. As in [7], the "*non-neoplastic*" and "*benign*" label is treated as the benign label while the "*malignant*" label is treated as the malignant label.

ImageNet-100. The images are resized to 224 pixels along the shorter side using bilinear interpolation. The training and validation splits are provided in https://github.com/Crazy-Jack/Cl-InfoNCE/tree/main/data_processing/imagenet100/hier. In the hierarchical label experiments, we select the labels from the 5-th and 6-th level of the WordNet hierarchy as the weak labels. The semantic grouping of the labels is given in Table A.1.

Level 5	Level 6
artifact	covering, instrumentality, structure, surface
food	fare, foodstuff, nutriment, produce
organism	animal, fungus, plant

Table A.1. Coarse-grained labels selected for the hierarchical label experiments.

B. Data Augmentation

The visual attributes (UT Zappos 50k, WIDER Attribute, CUB-200-2011) and hierarchical labels (ImageNet-100) experiments follow the standard augmentations which include random scaling, cropping, horizontal flipping, and color jittering. The fairness (Fitzpatrick17k) experiment uses random cropping, flipping, and rotation. The images are converted into tensors with values within the range of $[0, 1]$ and normalized using the dataset-specific mean and standard deviation (except for Fitzpatrick17k, which uses statistics computed from ImageNet). The augmentation parameters are given in Table B.1.

C. Hyperparameter Settings

Except for the pretraining step in Fitzpatrick17k, all remaining experiments were conducted using the same hyperparameter settings as those established in prior works [4, 7]. The specific hyperparameters used in the pretraining and downstream experiments are respectively tabulated in Table C.1 and Table C.2.

References

[1] Jia Deng, W. Dong, R. Socher, Li-Jia Li, K. Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, 2009. 2

[2] Matthew Groh, Caleb Harris, Luis Soenksen, Felix Lau, Rachel Han, Aerin Kim, Arash Koochek, and Omar Badri. Evaluating deep neural networks trained on clinical images in dermatology with the fitzpatrick 17k dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1820–1828, 2021. 2

[3] Yining Li, Chen Huang, Chen Change Loy, and Xiaoou Tang. Human attribute recognition by deep hierarchical contexts. In *European Conference on Computer Vision*, 2016. 2

[4] Yao-Hung Hubert Tsai, Tianqin Li, Weixin Liu, Peiyuan Liao, Ruslan Salakhutdinov, and Louis-Philippe Morency. Learning weakly-supervised contrastive representations. In *International Conference on Learning Representations*, 2022. 1

[5] C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie. The caltech-ucsd birds-200-2011 dataset. Technical Report CNS-TR-2011-001, California Institute of Technology, 2011. 2

[6] Aron Yu and Kristen Grauman. Fine-grained visual comparisons with local learning. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 192–199, 2014. 2

[7] Yongshuo Zong, Yongxin Yang, and Timothy Hospedales. MEDFAIR: Benchmarking fairness for medical imaging. In *International Conference on Learning Representations*, 2023. 1

Dataset	Access link
UT Zappos 50k [6]	https://vision.cs.utexas.edu/projects/finegrained/utzap50k/
WIDER Attribute [3]	http://mmlab.ie.cuhk.edu.hk/projects/WIDERAttribute.html
CUB-200-2011 [5]	http://www.vision.caltech.edu/datasets/cub_200_2011/
Fitzpatrick17k [2]	https://github.com/mattgroh/fitzpatrick17k
ImageNet-100 [1]	Original data: https://image-net.org/download.php Selected 100 classes: https://github.com/Crazy-Jack/Cl-InfoNCE

Table A.2. Access links to the datasets.

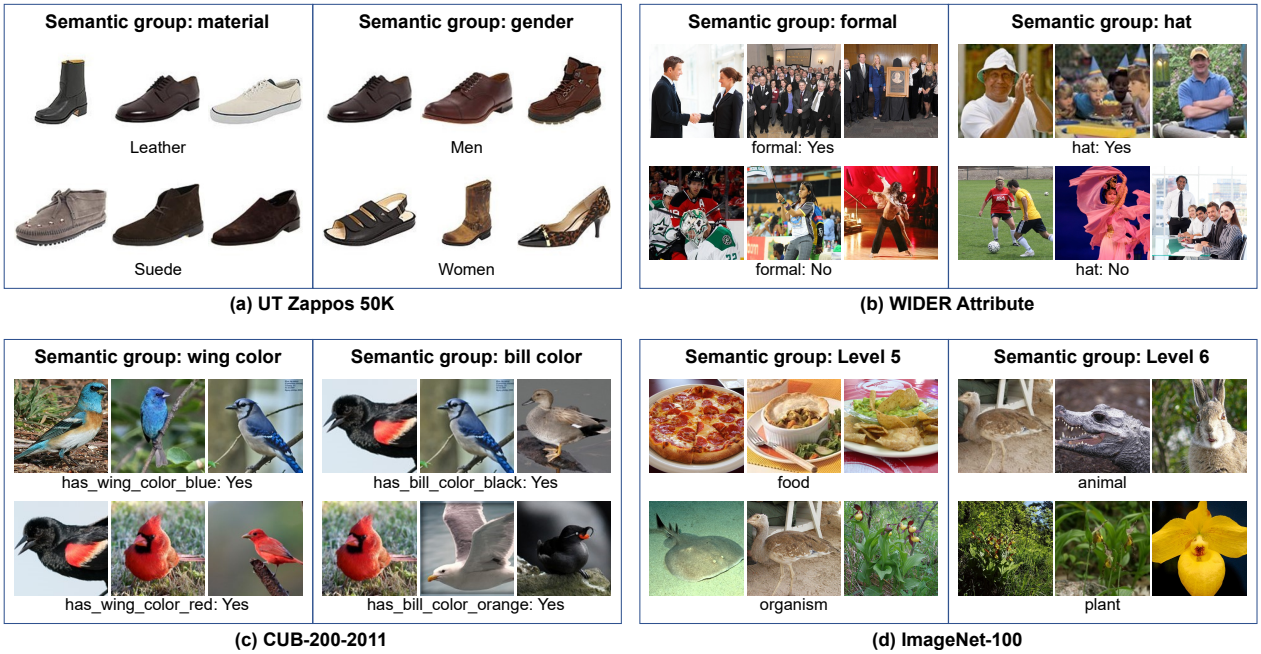

Figure A.1. Visualization of semantic groups for the benchmark datasets. Within each group, the images in the same row share the same visual attribute and form positive pairs with each other in CLRC.

Parameter	Attributes & Hierarchical	Fairness
Crop size	224 x 224 / 32 x 32	224 x 224
Scaling factor	$[0.3, 1.0]$	1.0
Horizontal flipping - probability	0.5	0.5
Color jittering - probability	0.8	-
Color jittering - brightness factor	$[0.6, 1.4]$	-
Color jittering - contrast factor	$[0.6, 1.4]$	-
Color jittering - saturation factor	$[0.6, 1.4]$	-
Color jittering - hue factor	$[-0.1, 0.1]$	-
Gray scaling - probability	0.2	-
Rotation - degree	-	$[-15, 15]$

Table B.1. Parameters of data augmentations.

Parameter	UT Zappos 50k	WIDER Attribtue	CUB-200-2011	ImageNet-100	Fitzpatrick17k
Batch size	128	40	128	128	512
Optimizer	SGD	SGD	SGD	SGD	SGD
Learning rate	0.17	0.1	0.1	0.1	0.01
Learning rate scheduler	cosine	cosine	cosine	cosine	cosine
Weight decay	$1e-4$	$1e-4$	$1e-4$	$1e-4$	$1e-4$
Max epochs	1000	1000	1000	200	50
Warmup epochs	330	330	330	66	15
Warmup scheduler	linear	linear	linear	linear	linear
Contrastive Temperature	0.07	0.1	0.1	0.1	0.1

Table C.1. Hyperparameters used in the pretraining experiments.

Parameter	UT Zappos 50k	WIDER Attribute	CUB-200-2011	ImageNet-100	Fitzpatrick17k
Type	linear	linear	linear	linear	fine-tune
Batch size	128	40	128	1024	1024
Optimizer	SGD	SGD	SGD	SGD	SGD
Learning rate	0.3	0.3	0.3	0.4	0.001
Learning rate scheduler	cosine	cosine	cosine	cosine	none
Weight decay	$1e-4$	$1e-4$	$1e-4$	0.0	$1e-5$
Max epochs	100	100	100	90	until early stop
Warmup epochs	33	33	33	0	0
Warmup scheduler	linear	linear	linear	none	none

Table C.2. Hyperparameters used in the downstream experiments.